**LAB\_WORKSHEET\_WEEK[作业标题]**

**4th Dec,2019[日期]**

1.作业题目：**[Assignment上复制一段]**

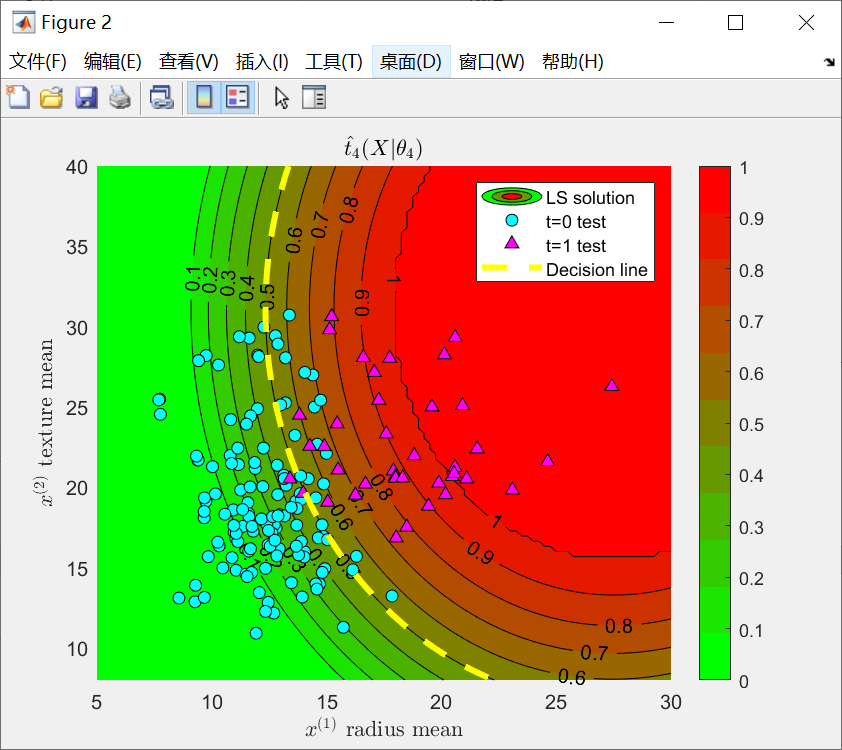
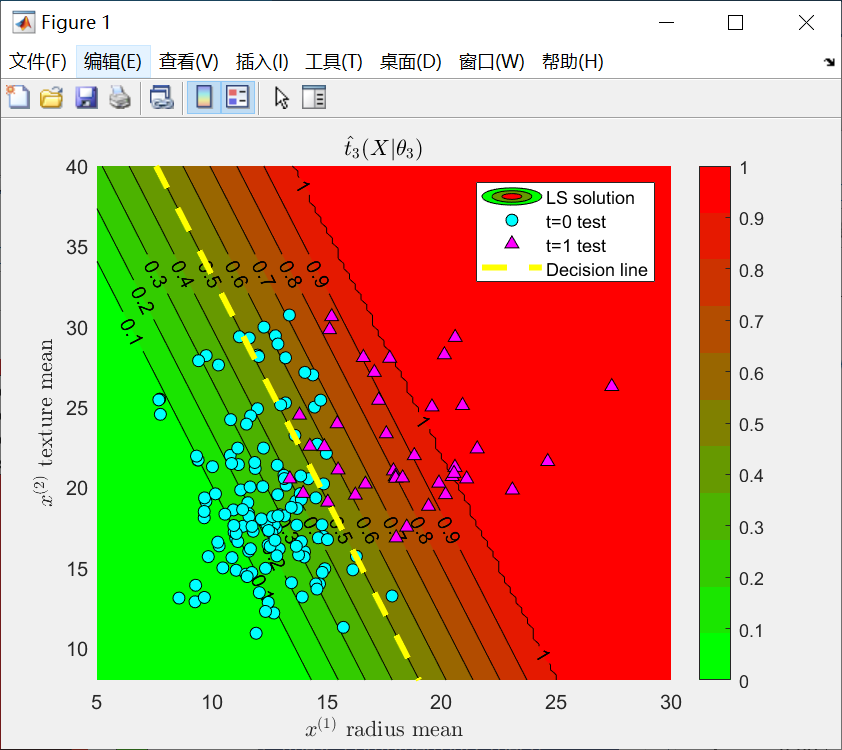
Follow the Instructions (Section 3 of this document) to fill in the details of the functions in the template file.

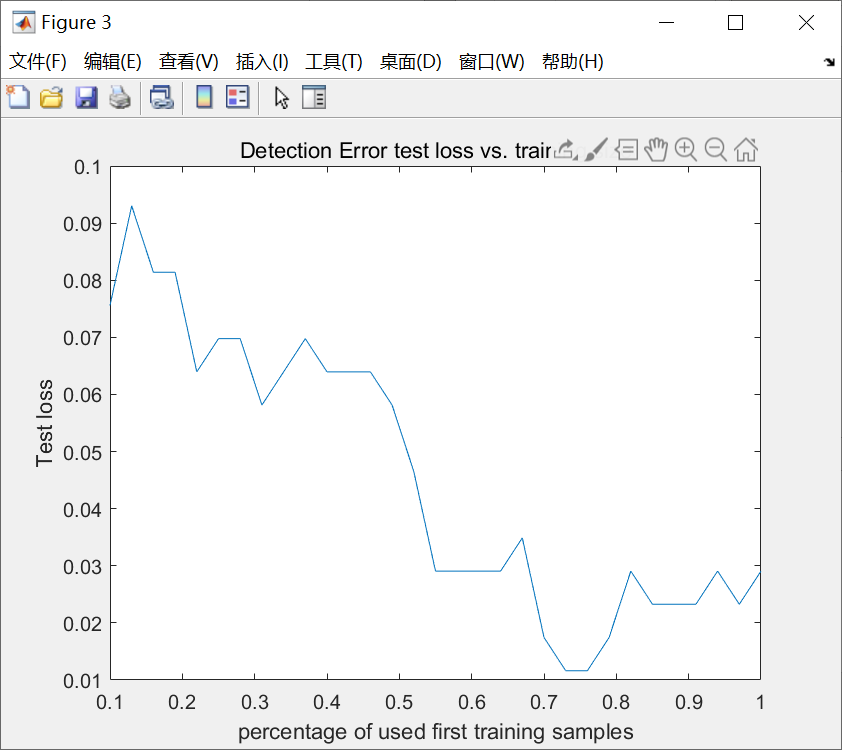
Once you have written the functions, verify k12345678.m runs without errors when the file is included in a folder containing only the file itself and the data set dataset\_wisconsin.mat.

2.作业目的：**[作业的目的1-3个点都可以]**

1. 复习巩固matlab的基础语法知识
2. 掌握机器学习的相关知识和matlab实现，以及模型结果的验证

3.运行效果： **[运行结果截图]**

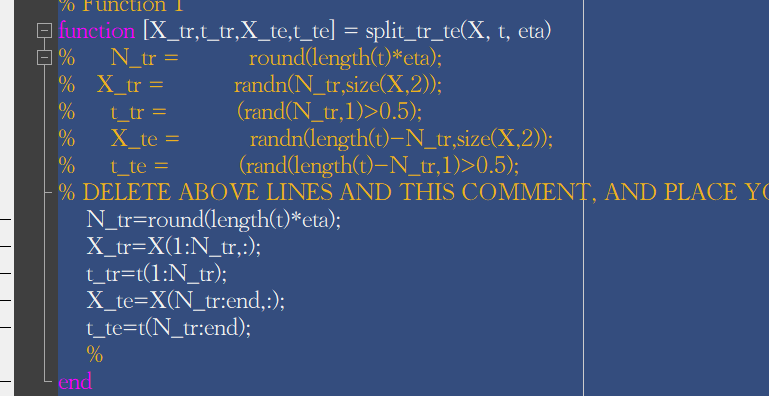




4.实现过程：**[部分核心代码截图, 截图部分的代码prefer 中英文注释]**

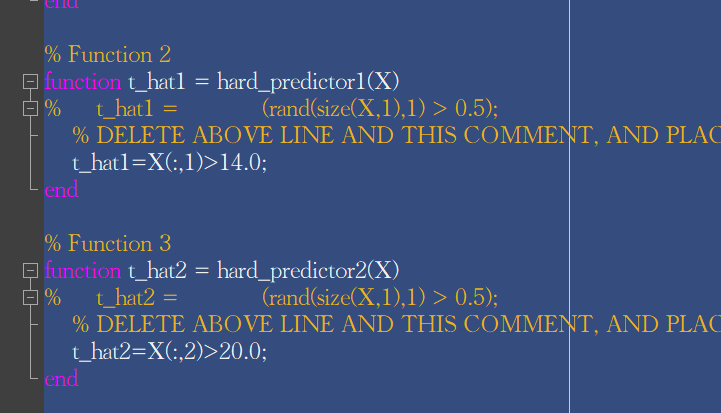
Section 1——split the data set

使用前个数据作为训练集，剩余的部分为测试集



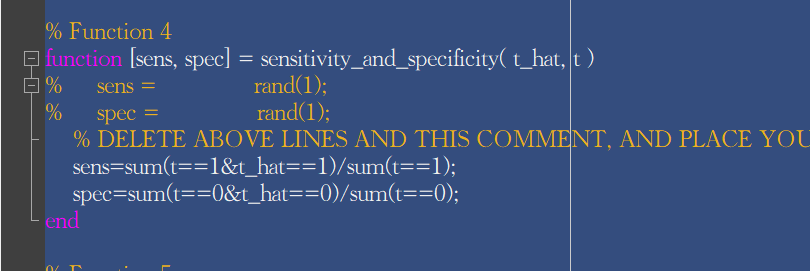
Section2——single-feature-based thresholding hard predictors

使用单一指标进行预测，分别使用第一个指标和第二个指标进行预测



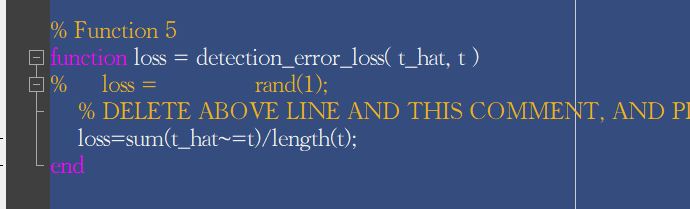
Section3——calculate the sensitivity and specificity

计算敏感性和特异性，根据分类预测结果，计算敏感性和特异性



Section4——estimate the detection error loss

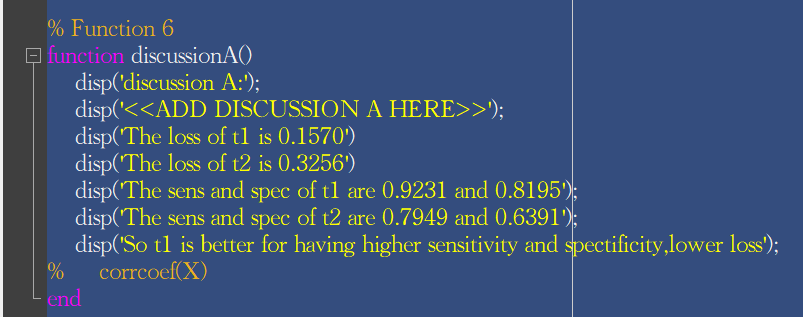
根据分类预测结果，计算经验误差



这里定义empirical detection error loss

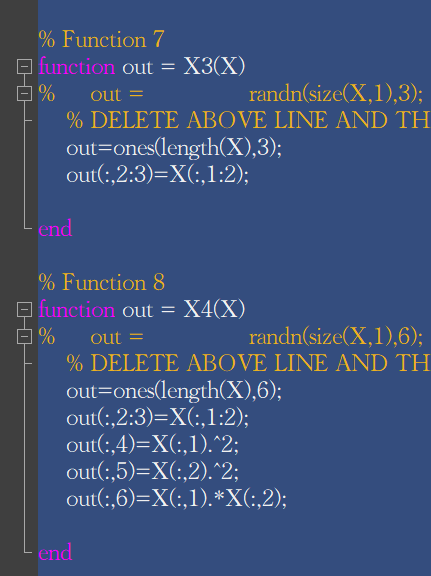
Section5——discussion A

根据之前的结果，进行讨论，判断两个指标哪个更优，经过讨论后，确定为指标



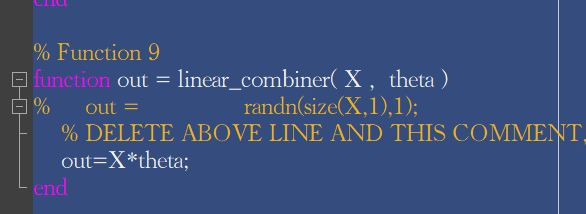
Section6——produces the data matrix

针对两种组合，利用源数据生成两个数值矩阵



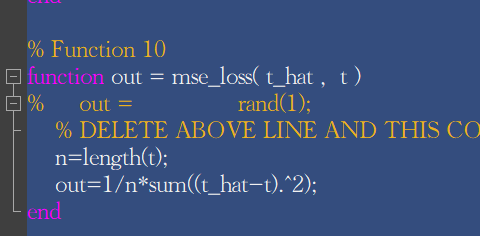
Section7——applies the predictor

得到数值矩阵和值后，相乘进行预测值输出



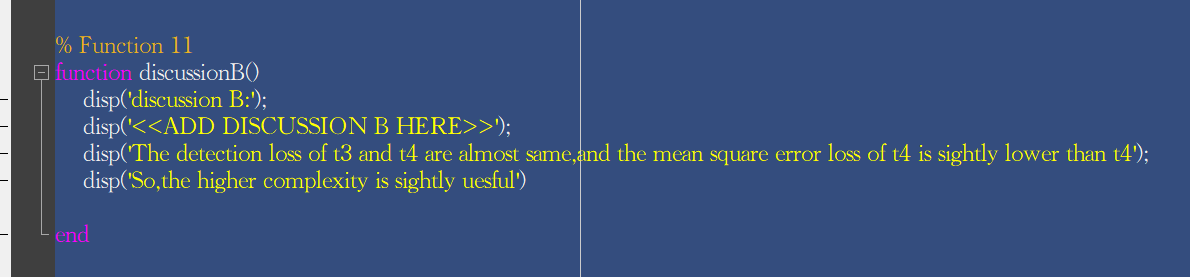
Section8——evaluate the mean square error loss

计算均方误差



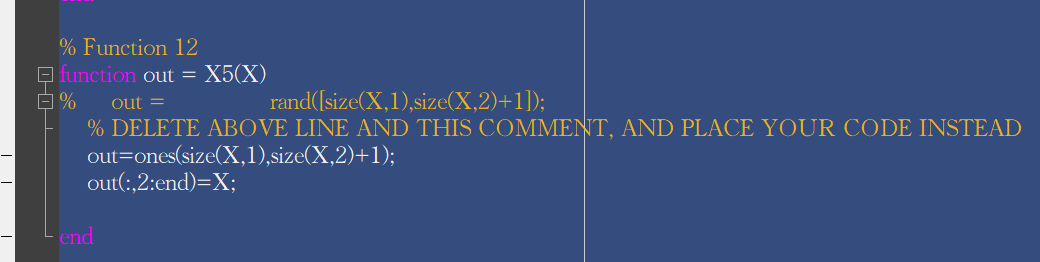
Section9——discussion B

根据之前的结果，进行评估，确定和选择



Section10——produces the data matrix

生成全特征的数值矩阵

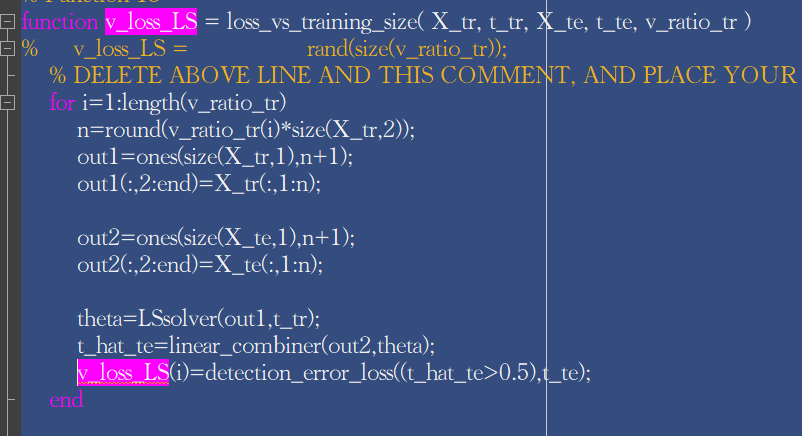


Section11——quantify how the full-features least square solution test loss is sensitive to

the amount of training data samples

探索loss与训练样本数量的敏感性

按照10:3:100划分训练样本数量，每次仿照之前的方法进行预测，并且绘制结果



参考源码图 1.1

5.知识点巩固：**[相关知识点总结]**

- **敏感性 Sensitivity**

敏感性又称真阳性率，就是发病之后，你的诊断方法对疾病的敏感程度（识别能力）。

敏感性越高，漏诊概率越低。

* **特异性 Specificity**

特异性又称真阴性率，不发病（我们这里称之为健康）的特征是有别于发病的特征的，我们利用这些差异避免误诊，那么诊断标准对于这些差异利用的如何就用特异性来表示。

特异性越高，确诊概率越高。

- **均方误差 mean square error loss**

6.知识拓展：**[拓展知识点1-3点总结]**

**- 过拟合**

过拟合（over-fitting）也称为过学习，它的直观表现是算法在训练集上表现好，但在测试集上表现不好，泛化性能差。过拟合是在模型参数拟合过程中由于训练数据包含抽样误差，在训练时复杂的模型将抽样误差也进行了拟合导致的

7.学习建议：**[根据学习IT的经验，写1-3点, prefer 3个点]**

1、遇到问题可以上百度，必应，但是如果想找到英文资料，最好还是Google

2、机器学习关键在于指标的选取和处理方法，这次lab中有一个很好的方法，就是提升指标次数，这个方法可以借鉴学习